README File for 2010 Census Summary File 2 Delivered via FTP

Note: *Users processing these FTP files in a Windows environment should read carefully the File Information section of this document.*

Contents

About the FTP Application

This FTP (File Transfer Protocol) application is intended for experienced users of census data, zip files, and spreadsheet/database software. It provides quick access for data users such as State Data Centers and news media needing to begin their analysis immediately upon data release. Due to the size of the files, the FTP user should have a fast file transfer capability. See Table 1. Estimated File Sizes on pages 3-5.

The FTP Directory

The FTP directory is at http://www2.census.gov/census_2010/05-Summary File_2. When the 2010 Census Summary File 2 are added to each state's directory, it will contain a single zip package with the geographic header file, 11 data segment files for each of the applicable 331 iterations totaling as many as 3,641 files, and a packing list which provides information about the file's creation and size. See below for more information on Segmented Data.

Users of the FTP application need to unzip the package after downloading, then import the data into the spreadsheet/database software of their choice for data analysis and table presentation. *We are unable to provide one-on-one support for applications of the data to specific spreadsheets or data base software.*

File Information

Once the package is unzipped, the files are in flat ASCII format. The geographic header file contains fixed fields while the data files are in comma-delimited format. These are text files however the file extension is not '.txt'. The user will need to rename the files with a .txt extension for import into some software packages, e.g. Microsoft Access. These files have been constructed

in a LINUX environment. They use an ASCII linefeed, chr(10), to indicate a new record.

For successful use with many programs running in a Windows environment, these files need to be modified to use the ASCII carriage return/linefeed sequence, chr(13) + chr(10) as a record terminator. This is an easy step in the UnZIP process using any UnZIP software which offers the conversion option. Winzip for Windows, version 14.5 has been tested.

The resulting file will meet the ANSI MS-DOS/Windows standard used by Access 2003 and Access 2007 and other MS Windows-based programs. If the data are being processed in a LINUX environment, they can be unzipped using any standard Linux ZIP/UnZIP package.

Note to Users of Microsoft Access:

Due to the FieldSize property limitations within Microsoft Access, modifications to field types are required when importing the Geographic Header Record file and File01 through File11:

- Fields classified as numeric (N) should be imported as long integers unless the field has been described as having decimals. These fields should be imported as double.
- AREALAND and AREAWATR should be imported as text.
- Also see Access database structure and instructions (HowToUseMSAccessWithSummaryFile2.pdf) for converting the data at http://www2.census.gov/census_2010/05-Summary_File_2.
- Failure to make these changes may result in missing data upon import.

Segmented Data

The data in the 2010 Census Summary File 2 and other 2010 Census summary files are segmented. This is done to manage the volume of data and to facilitate exporting into spreadsheet or database software. The data and the corresponding geographic information for an individual state are known as the file set. Because of the large size of the tables, the file set will be broken into as many as 3,643 files: a geographic header record file, 11 data segment files for each of the applicable 331 iterations, and a packing list. To get a complete set of the 2010 Census Summary File 2, users must download the geographic header file and all the data file segments in the package.

Table 2 below provides the file/table details. When using Table 2 you must take into consideration that the first 5 fields of the data file contains identification information (FILEID, STUSAB, CHARITER, CIFSN, and LOGRECNO).

Table 2. File/Table Segmentation

Data file segment number	Number of data cells	Starting matrix number	Ending matrix number
1	1	PCT1	PCT1
2	6	PCT2	PCT2
3	212	PCT3	PCT4
4	223	PCT5	PCT18
5	233	PCT19	PCT31
6	124	PCT32	PCT38
7	219	PCT39	PCT47
8	235	PC01	PC07
9	89	PC08	PC10
10	6	HCT1	HCT1
11	172	HCT2	HCT14

The explanation below for linking the two data files requires specific location information about the geographic header. These are located in Chapter 2-How to use this Product, of the Technical Documentation http://www.census.gov/prod/cen2010/doc/sf2.pdf.

A unique logical record number (LOGRECNO) in the geographic header is assigned to all records for a specific geographic entity. All records for that entity can be linked together across files. The geographic header **record layout** is identical across all electronic data products from the 2010 Census. However, the content is product specific. Some header fields that appear in both file types (geographic header and fileXX) are not used. For example, the characteristic iteration (CHARITER) field is used in the 2010 Census Summary File 2 data products but not in Summary File 1, it is always coded as 000.

File Record Layout

For a layout of the data table, see http://www.census.gov/prod/cen2010/doc/sf2.pdf, select Chapter 6, *Data Dictionary*.

Table 1. Estimated File Sizes

State	STUSAB	Zipped Estimate Mb
01	AL	92
02	AK	28
04	AZ	135
05	AR	73
06	CA	841
08	CO	106
09	CT	78
10	DE	20

11	DC	9
12	FL	343
13	GA	167
15	HI	58
16	ID	32
17	IL	265
18	IN	118
19	IA	76
20	KS	71
21	KY	72
22	LA	102
23	ME	26
24	MD	119
25	MA	132
26	MI	162
27	MN	112
28	MS	67
29	MO	122
30	MT	29
31	NE	45
32	NV	66
33	NH	24
34	NJ	171
35	NM	65
36	NY	371
37	NC	204
38	ND	18
39	OH	189
40	OK	153
41	OR	84
42	PA	199
44	RI	22
45	SC	85
46	SD	25
47	TN	117
48	TX	475
49	UT	56
50	VT	15
51	VA	159
53	WA	154

54	WV	36
55	WI	107
56	WY	16
72	PR	78
TOTAL		**6.3 Gb**